

课后作业三：支持向量机

任务：基于 SVM 构建机器学习模型，对 Omniglot 中的字符分类。

数据集：Omniglot Dataset 是全语言文字数据集，包含各种语言的不同字母表，如日语的平假名和片假名，韩语的元音和辅音，最常见的拉丁字母 abcd 等。Omniglot Dataset 共包含 50 个不同语言的字母表，每个字母表中包含不同的字符，共 1623 种字符，每种字符由 20 个不同的人书写。本次实验选取其中 200 种字符，每种字符的 15 张图片作为训练集，5 张图片作为测试集。

实现：

1. 导入必要的库

```
import csv
import numpy as np
import pandas as pd
import scipy.io
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
```

机器学习中非常重要的库，包括一些分类、回归、聚类、降维、模型选择和预处理

2. 加载数据

```
train_data = scipy.io.loadmat(r'D:\dataenclosure\third\train_data.mat')
test_data = scipy.io.loadmat(r'D:\dataenclosure\third\test_data.mat')
```

3. 获取训练数据和标签，并划分训练集和测试集。从 .mat 文件中加载训练数据并展平图像数据。生成对应的标签，每个类别有 15 张图片，共有 200 个类别。

```
X_train = train_data['train']
X_train = X_train.reshape(-1, 28 * 28) # 将图像数据展平为一维数组
y_train = np.repeat(np.arange(1, 201), 15) # 生成标签
# 划分训练集和验证集
X_train, X_val, y_train, y_val = train_test_split(X_train, y_train, test_size=0.2,
random_state=42)
```

4. 初始化和训练 SVM 模型，创建一个线性核的 SVM 模型，并用训练数据进行训练。

```
# 初始化 SVM 模型
svm_model = SVC(kernel='linear', C=1.0, random_state=42)
# 训练 SVM 模型
svm_model.fit(X_train, y_train)
```

5. 用训练好的模型在验证集和测试集上进行预测。

```
# 在验证集上进行预测
y_val_pred = svm_model.predict(X_val)
# 在测试集上进行预测
X_test = test_data['test']
X_test = X_test.reshape(-1, 28 * 28) # 将测试数据展平为一维数组
```

```
y_test_pred = svm_model.predict(X_test)
L = list(range(1, 1001))
```

6.将测试集的预测结果保存到 submission.csv 文件中

```
file_path = r'D:\dataenclosure\third\submission.csv'
with open(file_path, 'w', newline='', encoding='utf-8') as f:
    fieldnames = ['图片', '预测结果']
    f_csv = csv.DictWriter(f, fieldnames=fieldnames)
    f_csv.writeheader()
    for i in range(0, len(y_test_pred)):
        f_csv.writerow({'图片':L[i], '预测结果':y_test_pred[i]})
pass
```

7.结果如下:

只展示部分:

图片	预测结果	
1	8	
2	173	
3	195	
4	13	
5	116	
6	74	
7	189	
8	135	
9	15	
10	33	
11	27	
12	119	
13	171	
14	138	
15	122	
16	95	
17	117	
18	12	
19	184	
20	140	
21	72	
22	181	
23	13	
24	187	
25	178	
26	120	
27	22	
28	24	
29	18	